

AZBUKA I KODOVI

AZBUKU možemo predstaviti kao skup i operacije nad njim

- konačan neprazni skup proizvoljnih simbola - konačna azbuka V
- elementi skupa V - slova ili simboli
- konačne niske slova iz V - reči

- tada je jezik L – proizvoljan skup reči iz V^* , gde je V^* skup svih reči nad azbukom V (gde je λ prazna reč)

- dužina reči $|P|$ – broj slova u reči P . Važi $|\lambda| = ?$

- $V^+ = V^* \setminus \{\lambda\}$

Primeri azbuka i jezika nad njima

• Za azbuku $V_1 = \{a, b\}$ možemo da definišemo sledeće jezike:

– $L_1 = \{a, b, \lambda\}$ – jezik sa rečima dužine ≤ 1

– $L_2 = \{a_i b_i \mid i \in \{0, 1\}\}$ – jezik sa rečima dužine ≤ 2

– ...

• Za azbuku $V_2 = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ reči nad V_2^+ su brojevi u dekadnom brojevnom sistemu (sa eventualnim vodećim nulama)

• Za azbuku $V_3 = \{0, 1\}$ reči nad V_3^+ su brojevi u binarnom brojevnom sistemu (sa eventualnim vodećim nulama)

- KODOVE možemo predstaviti kao preslikavanja jezika nad azbukama

$$V_1 = \{ a_1, a_2, \dots, a_m \}, m \geq 1, \quad V_2 = \{ b_1, b_2, \dots, b_n \}, n \geq 2 \text{ azbuke}$$

$$L_1 \subset V_1^*, L_1 = \{ p_1, p_2, \dots, p_r \}, r \geq 1, \quad L_2 \subset V_2^*, L_2 = \{ q_1, q_2, \dots, q_s \}, s \geq 1 \text{ jezici nad njima}$$

1. *kodiranje* je svaka funkcija f definisana sa $f : L_1 \rightarrow L_2$.

2. *dekodiranje* je funkcija $g : L_2 \rightarrow L_1$, $g = f^{-1}$, ako f^{-1} postoji.

3. Kôd jezika L_1 u azbuci V_2 je skup vrednosti $K \subseteq L_2$, $K = \{ q_j \mid \exists_i : f(p_i) = q_j \}$

(Kodiranje je izračunavanje vrednosti $f(p_i)$. Dekodiranje je izračunavanje vrednosti $g(q_j)$.)

4. Kôd je jednoznačan akko je funkcija f 1-1. U suprotnom je kod višeznačan.

5. Broj simbola u azbuci V_2 se naziva osnova koda. Reč q_i iz jezika L_2 se naziva kodna reč.

6. Kôd je potpun kada obuhvata sve reči određene dužine u jeziku L_2 .

7. Kôd je ravnomeran ako $\exists_1 \forall_j : |q_j| = 1$, tj. ako je dužina svih kodnih reči u jeziku L_2 ista. U suprotnom, kôd je neravnomeran. Da bi kôd dužine d reči P bio ravnomeran mora da važi $|P| \leq n^d$ gde je n broj simbola u azbuci V_2 . Ako važi $|P| = n^d$ tada je u pitanju potpun ravnomeran kod. Dužina reči ravnomernog koda se naziva mesnost koda.

Primeri:

* Neka azbuka $V_1 = \{ +, -, *, / \}$ sadrži oznake elementarnih aritmetičkih operacija. Za kodiranje odgovarajućeg jezika $L_1 = \{ +, -, *, / \}$ u binarnoj azbuci dovoljno je uzeti reči dužine 2. Neka moguća kodiranja su prikazana u narednoj tabeli. Svi prikazani kodovi su ravnomerni i potpuni.

Reč u jeziku	Kod 1	Kod 2	Kod 3	Kod 4	Kod 5	Kod 6
+	00	00	00	00	00	00
-	01	01	10	10	11	11
*	10	11	01	11	01	10
/	11	10	11	01	10	01

* Ukoliko želimo da zapišemo reči živog jezika tada za azbuku V_1 moramo da uzmemo sve znake pisma (velika i mala slova, cifre, interpunkcijske i specijalne znake). Dužina kodnih reči će zavisiti od broja simbola u azbuci, odnosno broja reči u odgovarajućem (formalnom) jeziku L_1 . U ranijem periodu razvoja računarstva postojalo je više različitih kodova, ali se danas koriste kodovi sa dužinom reči 7, 8 ili 16.

ZAPIS PODATAKA

Podaci:

- znakovni (alfanumerički)
- podskup brojčanih (numeričkih) podataka u koje spadaju
 - celi brojevi
 - binarno kodirani dekadni brojevi
 - realni brojevi zapisani u nepokretnom zarezu
 - realni brojevi zapisani u pokretnom zarezu.

ZAPIS ZNAKOVNIH PODATAKA

- Znakovni podaci se u računaru zapisuju pomoću binarnih kodova.
 - Koristi se više različitih kodova, kod kojih se svaki karakter skupa alfanumeričkih podataka kodira niskom binarnih cifara dužine 7, 8 ili 16.
-

1. ASCII (American National Standard Code for Information Interchange). 7-bitni kod, 128 karaktera. Osmi bit se koristi za kontrolu parnosti.

2. EBCDIC(ExtendedBinaryCodedDecimalInterchangeCode). 8-bitni kod, 256 karaktera.

3. ISO-8 (International Organization for Standardization-8). 8-bitni kod, 256 karaktera. Prvih 127 pozicija poklapa se sa ASCII kodom. Pozicije iznad 126 su popunjene različitim kontrolnim i grafičkim

4. IBM-PC kod. 8-bitni 8-bitni kod, 256 karaktera. Prvih 127 pozicija poklapa se sa ISO-8 kodom.

5. UNICODE (UNIversal encODE). 16-bitni kod, 65536 karaktera.

6. UTF-8 (Unicode Transformation Format). UTF-8 kodira svaki UNICODE simbol pomoću promenljivog broja (1-4) okteta (svaki oktet ima 8 bitova), pri čemu broj okteta zavisi od celobrojne vrednosti koja je dodeljena UNICODE karakteru. UTF-8 je kompatibilan (naniže) sa ASCII kodom. Danas je u širokoj upotrebi.

- Kodna strana (eng. code page). predstavlja preslikavanje između grafičkih karaktera i skupa heksadekadnih identifikacija.

Primer JACA

1. ASCII - 0100 1010 0100 0001 0100 0011 0100 0001
2. EBCDIC - 1101 0001 1100 0001 1100 0011 1100 0001
3. ISO-8 - 0100 1010 0100 0001 0100 0011 0100 0001
4. UNICODE – U+004A U+0041 U+0043 U+0041

HTML code: J A C A

[http:// unicode-table.com/en/](http://unicode-table.com/en/)

5. UTF 8 - 0100 1010 0100 0001 0100 0011 0100 0001

<http://dev.networkerror.org/utf8/>

<http://www.fileformat.info/info/unicode/char/004a/index.htm>

Domaći zadatak:
Pronaći zanimljivosti o kodovima
(npr šta je YUSCII)